Modeling and Estimating Uncertainty in Parameter Estimation

H.T. Banks and Kathleen L. Bihari
Center for Research in Scientific Computation
Box 8205
North Carolina State University
Raleigh, NC 27695-8205
Fax: 919 515-1636
email: htbanks@eos.ncsu.edu

Abstract

In this paper we discuss questions related to reliability or variability of estimated parameters in deterministic least squares problems. By viewing the parameters for the inverse problem as realizations for a random variable we are able to use standard results from probability theory to formulate a tractable probabilistic framework to treat this uncertainty. We discuss method stability and approximate problems and are able to show convergence of solutions of the approximate problems to those of the original problem. The efficacy of our approach is demonstrated in numerical examples involving estimation of constant parameters in differential equations.

1

1. REPORT DATE **1999**	2. REPORT TYPE	3. DATES COVERED **00-00-1999 to 00-00-1999**	
4. TITLE AND SUBTITLE **Modeling and Estimating Uncertainty in Parameter Estimation**		5a. CONTRACT NUMBER	
		5b. GRANT NUMBER	
		5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)		5d. PROJECT NUMBER	
		5e. TASK NUMBER	
		5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **North Carolina State University,Center for Research in Scientific Computation,Raleigh,NC,27695-8205**		8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)	
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**			
13. SUPPLEMENTARY NOTES **The original document contains color images.**			
14. ABSTRACT **see report**			
15. SUBJECT TERMS			

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES **23**	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**			

1   Introduction

A standard deterministic inverse problem frequently encountered in both applied and theoretical literature can be abstractly stated as follows: Given a parameter dependent dynamical or algebraic system

$$\mathcal{A}(q, u(q)) = \mathcal{F}(q) \tag{1}$$

with states $u(q)$, parameters q and operators (differential or algebraic) $\mathcal{A}$, use observations (possibly incomplete) or data on the states to determine the best parameters q^* in some admissible set Q so that the solution of equation (1) for $q = q^*$ best describes the data. For such deterministic problems there is a large literature based on diverse formulations (least squares, equation error, etc.). For discussions of some of these see [3]. Once one has "solved" this (by no means trivial) deterministic problem, it is frequently important to know something about the reliability of the estimates. One approach entails attaching "error bars" to the estimated parameter values, much like one does in standard statistical analysis or in scientific computational analysis (using *a priori* bounds) with finite discretization techniques (finite difference, finite elements, etc) from numerical analysis. In essence we are asking for measurements of *uncertainty* (inherent in our methods rather than in our data collection) related to our best estimates of parameters q. Thus we are led in a completely natural way to stochastic or probabilistic aspects of estimates from a deterministic problem solved with deterministic algorithms.

We offer here ideas for one approach to treatment of variability in parameter estimation techniques. The approach is based on viewing multiple observations $\{\hat{u}_j\}_{j=1}^N$ of the state in (1) as observations corresponding to a set of realizations $\{q_j\}_{j=1}^N$ of the parameter q which is now thought of as a specific (albeit unknown) random variable with probability distribution P on Q. The system (1) is accordingly reformulated in terms of the state $u = u(P)$ depending on the probability distribution P. The observations $\{\hat{u}_j\}$ can be averaged so that $\hat{u} = \frac{1}{N} \sum_{j=1}^N \hat{u}_j$ is an observation for $u(P)$ and one can then attempt to estimate a best distribution P^* to fit this data $\hat{u}$, for example in some type of least squares fit. Once one obtains P^*, its mean μ and variance σ^2 can be used as a best parameter estimate and measure of reliability, respectively.

In the sections below we give a concrete example of this (using nonlinear parameter dependent ordinary differential equations for the system (1)). We present a precise formulation of this conceptual approach, show that fundamental results from probability theory can be used to develop well-posedness results (existence, continuous dependence, and method stability) along with approximation ideas that are computationally tractable. We demonstrate feasibility of the resulting algorithms by presenting a summary of numerical findings using an example arising in estimation of effectiveness of vaccination policies in a population of susceptibles in disease prophylactics.

We believe the underlying philosophy as well as the specific formulation are applicable to and will be useful in a wide class of practical applications.

2 Parameter Estimation in Nonlinear Systems

To demonstrate our ideas, we will examine the estimation of constant parameters in a system of ordinary differential equations. These ideas can be easily extended to many systems of interest in applications, including systems of partial differential equations with unknown functional parameters (e.g., time and/or spatially dependent coefficients).

A typical estimation problem employs observations $\hat{x} = \{\hat{x_i}\}_{i=1}^n$ for $x(t_i)$, $i = 1, 2, \ldots, n$, to estimate parameters $q \in \mathbb{R}^m$ in the vector dynamical system

$$\dot{x}(t) = f(t, x(t), q). \tag{2}$$

Often a least squares formulation is used to find a best parameter value q^* in some admissible parameter set $Q \subset \mathbb{R}^m$. In other words, we attempt to find $q^* \in Q$ which is a minimum for

$$J(q, \hat{x}) = \sum_{i=1}^n |x(t_i; q) - \hat{x}_i|^2$$

over $q \in Q$ where $x(t_i; q)$ is a solution of (2) for a given $q \in Q$.

To introduce uncertainty, we view the parameters as realizations for a random variable and use the data to estimate the *probability distribution function* (PDF) for this random variable. Specifically, let $\mathcal{P}(Q)$ denote the set of probability distributions on Q and treat the data $\{\hat{x_i}\}$ as observations for the expected value

$$\mathcal{E}[x(t_i; q)|P] = \int_Q x(t_i; q) dP(q) \tag{3}$$

for a given PDF $P \in \mathcal{P}(Q)$. Note that if P is a discrete PDF with atoms $\{q_j\}_{j=1}^M \subset Q$ and associated probabilities $\{p_j\}$, $p_j \geq 0$, $\sum_{j=1}^M p_j = 1$, then (3) can be written

$$\int_Q x(t_i; q) dP(q) = \sum_{j=1}^M x(t_i; q_j) p_j.$$

Regardless of the form of P, the least squares estimation problem can be described as finding $P^* \in \mathcal{P}(Q)$ to minimize

$$J(P) = \sum_{i=1}^n |\mathcal{E}[x(t_i; q)|P] - \hat{x_i}|^2 \tag{4}$$

2

over $P \in \mathcal{P}(Q)$. To develop theoretical and computational results for this problem, it is necessary to have a topology on $\mathcal{P}(Q)$, continuity of the function $P \to J(P)$ in this topology, compatible compactness results, and some approximation results leading to implementable computational algorithms. In order to address these issues we will introduce the *Prohorov metric* and summarize some results from Billingsley [5].

3 The Prohorov Metric in the Space of Probability Distributions

Let $\mathcal{P}(Q)$ be the set of probability measures on the Borel subsets of Q, where Q is any complete metric space with metric d. For any closed subset $F \subset Q$ and $\epsilon > 0$, we define an ϵ-neighborhood of F by

$$F^\epsilon = \{q \in Q : d(\tilde{q}, q) < \epsilon, \ \tilde{q} \in F\}.$$

We then define $\rho : \mathcal{P}(Q) \times \mathcal{P}(Q) \to \mathbb{R}^+$ by

$$\rho(P_1, P_2) \equiv \inf\{\epsilon > 0 : P_1[F] \le P_2[F^\epsilon] + \epsilon, \ F \text{ closed}, \ F \subset Q\}.$$

The following properties of ρ are well-known:

(a) ρ is a metric (called the Prohorov metric) on $\mathcal{P}(Q)$;

(b) $(\mathcal{P}(Q), \rho)$ is a complete metric space;

(c) if Q is compact, then $(\mathcal{P}(Q), \rho)$ is a compact metric space.

We would like to understand convergence of $P_k \to P$ in the ρ metric. Unfortunately, the Prohorov metric is neither intuitive nor easy to use directly. However, it is well known that if (Q, d) is a complete metric space and $(\mathcal{P}(Q), \rho)$ is defined as above, then for $P_k, P \in \mathcal{P}(Q)$, the following convergence statements are equivalent:

(i) $\rho(P_k, P) \to 0$;

(ii) $\int_Q f dP_k(q) \to \int_Q f dP(q)$ for all bounded, uniformly continuous $f : Q \to \mathbb{R}^1$;

(iii) $P_k[A] \to P[A]$ for all Borel sets $A \subset Q$ with $P[\partial A] = 0$, where ∂A denotes the boundary of A.

The equivalence of (i) and (iii) reveals that convergence in the ρ metric is equivalent to convergence in distribution. Moreover, if we consider $\mathcal{P}(Q) \subset C_B^*(Q)$ where $C_B(Q)$ denotes the space of bounded, continuous functions on Q with the supremum norm, then (i) and (ii) imply that convergence in the ρ topology is equivalent

to weak* convergence in $\mathcal{P}(Q)$. For our discussions, we will make critical use of the equivalence between $\rho(P_k, P) \to 0$ and

$$\int_Q x(t;q)dP_k(q) \to \int_Q x(t;q)dP(q),$$

which in turn is the same as

$$\mathcal{E}[x(t;q)|P_k] \to \mathcal{E}[x(t;q)|P].$$

This convergence is needed to establish continuity in the ρ topology of the map

$$P \to J(P) = \sum_{i=1}^{n} |\mathcal{E}[x(t_i;q)|P] - \hat{x}_i|^2.$$

Continuity of this map, along with the compactness of Q, which guarantees compactness of $\mathcal{P}(Q)$ in the ρ metric, is sufficient to establish existence of a solution to the problem of minimizing (4) over $\mathcal{P}(Q)$.

If we assume existence questions are answered, and turn to the task of characterizing and/or finding minimizers, we note that $\mathcal{P}(Q)$ with the ρ metric is in general an infinite dimensional space because in general Q will be infinite dimensional. Thus, to address computational issues one must consider approximation ideas. To do this we first prove a density theorem that will be useful in establishing continuous dependence of estimates on data as well as in constructing approximation schemes.

There are numerous topologies on $\mathcal{P}(Q)$. If we define a $\mathcal{W}$-neighborhood of P as $N_\delta(P) = \{P_1 \in \mathcal{P}(Q) : P_1(F_i) < P(F_i) + \delta, \ i = 1, \cdots, k, \ F_i \text{ closed}, \ F_i \in \mathbb{S}\}$ for a given $\delta > 0$ and finite set $\{F_i\}_{i=1}^{k}$, this induces a topology on $\mathcal{P}(Q)$ which is equivalent to the topology of weak convergence, $\mathcal{W}$ (see [5], p. 236). We can also define a ρ-neighborhood of P by $\overline{N}_\epsilon(P) = \{P_1 \in \mathcal{P}(Q) : \rho(P, P_1) < \epsilon\}$ for a given $\epsilon > 0$. If Q is a separable space, the $\mathcal{W}$ topology is equivalent to the topology induced by the ρ-neighborhoods. We will be using the equivalence of these two topologies in the proof of Theorem 3.1. Here $\mathbb{N}^+$ are the positive integers, $\mathcal{R}$ are the rational numbers, and δ_{q_j} is the Dirac measure with atom at q_j.

Theorem 3.1 *Let Q be a complete, separable metric space with metric d, $\mathbb{S}$ be the class of all Borel subsets of Q and $\mathcal{P}(Q)$ be the space of probability measures on $(Q, \mathbb{S})$. Let $Q_0 = \{q_j\}_{j=1}^{\infty}$ be a countable, dense subset of Q. Then the set of $P \in \mathcal{P}(Q)$ such that P has finite support in Q_0 and rational masses is dense in $\mathcal{P}(Q)$ in the ρ metric. That is,*

$$\mathcal{P}_0(Q) \equiv \{P \in \mathcal{P}(Q) : P = \sum_{j=1}^{k} p_j \delta_{q_j}, \ k \in \mathbb{N}^+, \ q_j \in Q_0, \ p_j \in \mathcal{R}, \ p_j \geq 0, \ \sum_{j=1}^{k} p_j = 1\}$$

is dense in $\mathcal{P}(Q)$ relative to ρ.

4

<u>Proof</u>: Let $\epsilon > 0$ and let $P \in \mathcal{P}(Q)$. Let $\overline{N}_\epsilon(P)$ be a ρ-neighborhood of P. Since Q is separable, the $\mathcal{W}$ and ρ topologies are equivalent. Thus there is a $\delta > 0$ such that $N_\delta(P) \subseteq \overline{N}_\epsilon(P)$, where $N_\delta(P)$ is a $\mathcal{W}$-neighborhood of P of the form described above with closed sets $F_1, \cdots, F_k$.

Let $\{B_i\}_{i=1}^M$ be the partition of $\bigcup_{i=1}^k F_i \subset Q$ generated by the closed sets $F_1, \cdots, F_k$. We will assume each B_i is non-empty and so $M < \infty$. Since Q_0 is a dense subset of Q, $B_i \cap Q_0 \neq \emptyset$, for $i = 1, \cdots, M$.

For $i = 1, \cdots, M$, select a point $x_i \in B_i \cap Q_0$. At each point, x_i, place a mass, b_i, which satisfies the following three conditions: i) $b_i \in \mathcal{R}$, ii) $0 \leq b_i \leq P(B_i)$, and iii) $|P(B_i) - b_i| < \frac{\delta}{2M}$.

Now if $\cup_{i=1}^k F_i \neq Q$, select a point x_{M+1} so that $x_{M+1} \in Q_0 \cap (\cup_{i=1}^k F_i)^C$. If $\cup_{i=1}^k F_i = Q$, choose x_{M+1} so that $x_{M+1} \in Q_0 \setminus (\{x_i\}_{i=1}^M)$. In either case, place at x_{M+1} a mass $b_{M+1} \equiv 1 - \sum_{i=1}^M b_i$. Note that $b_{M+1} \in \mathcal{R}$ and $0 \leq b_{M+1} \leq 1$.

Define $P^* = \sum_{i=1}^{M+1} b_i \delta_{q_i}$. Then $P^*(Q) = \sum_{i=1}^{M+1} b_i = 1$, and $0 \leq P^*(A) \leq 1$ for all $A \in \mathbb{S}$. Thus $P^* \in \mathcal{P}_0(Q)$.

Define $K_i = \{j : F_i \cap B_j \neq \emptyset, 1 \leq j \leq M\}$. Note the set K_i has at most M indices.

Now suppose $\cup_{i=1}^k F_i \neq Q$. Then for any F_i,

$$
\begin{aligned}
|P^*(F_i) - P(F_i)| &= |P^*(\bigcup_{j \in K_i} B_j) - P(\bigcup_{j \in K_i} B_j)| \\
&= |\sum_{j \in K_i} P^*(B_j) - \sum_{j \in K_i} P(B_j)| \\
&= |\sum_{j \in K_i} b_j - \sum_{j \in K_i} P(B_j)| \\
&= |\sum_{j \in K_i} [b_j - P(B_j)]| \\
&\leq \sum_{j \in K_i} |b_j - P(B_j)| \\
&< \sum_{j \in K_i} \frac{\delta}{2M} \\
&\leq \frac{\delta}{2} \\
&< \delta.
\end{aligned}
$$

Now suppose $\cup_{i=1}^k F_i = Q$. If $x_{M+1} \notin F_i$, the above argument shows $|P^*(F_i) - P(F_i)| < \delta$. If $x_{M+1} \in F_i$, then

$$
\begin{aligned}
|P^*(F_i) - P(F_i)| \;&=\; |P^*(\bigcup_{j \in K_i} B_j) - P(\bigcup_{j \in K_i} B_j)| \\[2mm]
&=\; |\sum_{j \in K_i} P^*(B_j) - \sum_{j \in K_i} P(B_j)| \\[2mm]
&=\; |\sum_{j \in K_i} b_j + b_{M+1} - \sum_{j \in K_i} P(B_j)| \\[2mm]
&=\; |\sum_{j \in K_i} [b_j - P(B_j)] + [1 - \sum_{j=1}^{M} b_j]| \\[2mm]
&=\; |\sum_{j \in K_i} [b_j - P(B_j)] + [\sum_{j=1}^{M} P(B_j) - \sum_{j=1}^{M} b_j]| \\[2mm]
&=\; |\sum_{j \in K_i} [b_j - P(B_j)] + \sum_{j=1}^{M} [P(B_j) - b_j]| \\[2mm]
&\leq\; \sum_{j \in K_i} |b_j - P(B_j)| + \sum_{j=1}^{M} |P(B_j) - b_j| \\[2mm]
&<\; \sum_{j \in K_i} \frac{\delta}{2M} + \sum_{j=1}^{M} \frac{\delta}{2M} \\[2mm]
&\leq\; \frac{\delta}{2} + \frac{\delta}{2} \\[2mm]
&\leq\; \delta.
\end{aligned}
$$

Thus for all F_i, $P^*(F_i) < P(F_i) + \delta$, so $P^* \in N_\delta(P)$.

Since $N_\delta(P) \subseteq \overline{N}_\epsilon(P)$, $P^* \in \overline{N}_\epsilon(P)$. By construction $P^* \in \mathcal{P}_0(Q)$, so $\mathcal{P}_0(Q)$ is dense in $\mathcal{P}(Q)$ relative to ρ. $\square$

Theorem 3.1 can be used as a basis for defining a class of approximating sets to be used in tractable computational methods for the inverse problems defined in Section 2. First define

$$
Q_d = \bigcup_{M=1}^{\infty} Q_M \tag{5}
$$

6

where $Q_M = \{q_j^M\}_{j=1}^M$, $M = 1, 2, \cdots$, are chosen so that Q_d is dense in Q. Note that Q_d is countable. For each positive integer M let

$$\mathcal{P}^M(Q) = \{P \in \mathcal{P}(Q) : P = \sum_{j=1}^M p_j \, \delta_{q_j^M}, \ q_j^M \in Q_M, \ p_j \in \mathcal{R}, \ p_j \geq 0, \ \sum_{j=1}^M p_j = 1\}. \tag{6}$$

If we then define

$$\mathcal{P}_d(Q) = \cup_{M=1}^\infty \mathcal{P}^M(Q), \tag{7}$$

then by Theorem 3.1 we know $\mathcal{P}_d(Q)$ is dense in $\mathcal{P}(Q)$, and so we can approximate any element $P \in \mathcal{P}(Q)$ by a sequence $\{P_{M_j}\}$, $P_{M_j} \in \mathcal{P}^{M_j}(Q)$, such that $\rho(P_{M_j}, P) \to 0$ as $M_j \to \infty$.

4 Stability of the Inverse Problem

We now turn to the study of the inverse problem. We return to our original problem of finding a solution to

$$\min_{P \in \mathcal{P}(Q)} J(P, \hat{x}) = \sum_{i=1}^n |x(t_i, P) - \hat{x}_i|^2. \tag{8}$$

Given data $\hat{x}^k$ and $\hat{x}$ such that $\hat{x}^k \to \hat{x}$ as $k \to \infty$ and corresponding solutions $P^*(\hat{x}^k)$ and $P^*(\hat{x})$ (which in general are sets because there is not necessarily a unique minimizer of (8)), we say the problem is *continuously dependent on the data* (or *stable*) if $dist(P^*(\hat{x}^k), P^*(\hat{x})) \to 0$ as $k \to \infty$ (see [3, 4] for detailed discussions and motivation).

We now define a series of approximate problems. Let $\mathcal{P}^M(Q)$ be defined as in (6) where Q_d is a countable dense subset of Q as defined in (5) with $Q_M = \{q_j^M\}$. We define the approximate problem as finding a solution to

$$\min_{P_M \in \mathcal{P}^M(Q)} J(P_M, \hat{x}) = \sum_{i=1}^n |x(t_i, P_M) - \hat{x}_i|^2. \tag{9}$$

Let $P_M^*(\hat{x})$ denote the set of solutions for a given $\hat{x}$. The problems are *method stable* (again, see [3, 4] for further discussions) if for any data $\hat{x}^k$ and $\hat{x}$ such that $\hat{x}^k \to \hat{x}$ as $k \to \infty$ we have $dist(P_M^*(\hat{x}^k), P^*(\hat{x})) \to 0$ as $k \to \infty$ and $M \to \infty$. Note that this is equivalent to requiring $dist(P_M^*(\hat{x}^k), P_M^*(\hat{x})) \to 0$ as $k \to \infty$ uniformly in M.

Theorem 4.1 *Let Q be a compact metric space and assume solutions $x(t; q)$ of (2) are continuous in q on Q. Let $\mathcal{P}(Q)$ be the set of all probability measures on Q and let Q_d be a countable dense subset of Q as defined in (5) with $Q_M = \{q_j^M\}_{j=1}^M$. Define $\mathcal{P}_d(Q)$ as in (7) where $\mathcal{P}^M(Q)$ is defined as in (6). Suppose $P_M^*(\hat{x}^k)$ is the set*

7

of minimizers for $J(P)$ over $P \in \mathcal{P}^M(Q)$ corresponding to the data $\{\hat{x}^k\}$ and $P^*(\hat{x})$ is the set of minimizers over $P \in \mathcal{P}(Q)$ corresponding to $\{\hat{x}\}$ where $\hat{x}^k$, $\hat{x} \in \mathbb{R}^n$ are the observed data such that $\hat{x}^k \to \hat{x}$. Then $dist(P_M^*(\hat{x}^k), P^*(\hat{x})) \to 0$ as $M \to \infty$ and $\hat{x}^k \to \hat{x}$. Thus the solutions depend continuously on the data and the approximate problems are method stable.

<u>Proof</u>: Since Q is a compact, separable metric space, Q_d is dense in Q and $\mathcal{P}_d(Q)$ is the space of all probability measures with finite support in Q_d and rational masses, it follows from Theorem 3.1 that $\mathcal{P}_d(Q)$ is a dense subset of $\mathcal{P}(Q)$.

Since Q is compact, $(\mathcal{P}(Q), \rho)$ is compact, where ρ is the Prohorov metric. Since $q \to x(t, q)$ is continuous from Q to $\mathbb{R}^n$, whenever $P_M \to P$ in $\mathcal{P}(Q)$ we have that

$$
\begin{aligned}
\lim_{M \to \infty} J(P_M) &= \lim_{M \to \infty} \sum_{i=1}^{n} |\mathcal{E}[x(t_i, q)|P_M] - \hat{x}_i|^2 \\
&= \sum_{i=1}^{n} |\mathcal{E}[x(t_i, q)|P] - \hat{x}_i|^2 \\
&= J(P).
\end{aligned}
\tag{10}
$$

Since $\mathcal{P}^M(Q)$ is a closed subset of $\mathcal{P}(Q)$, $\mathcal{P}^M(Q)$ is compact in the ρ topology. Moreover, J is a continuous function on the compact set $\mathcal{P}^M(Q)$ and thus for a given set of observations $\hat{x}^k = \{\hat{x}_i^k\}_{i=1}^{n} \in \mathbb{R}^n$, there exists a (not necessarily unique) minimizer P_M^{*k} which is a solution to the problem of minimizing $J^k(P)$ over $P \in \mathcal{P}^M(Q)$ where

$$
J^k(P) \equiv J(P, \hat{x}^k) \equiv \sum_{i=1}^{n} |\mathcal{E}[x(t_i, q)|P] - \hat{x}_i^k|^2.
$$

Let $\{\hat{x}^k\}$ be a sequence that converges to some arbitrary $\hat{x} \in \mathbb{R}^n$. Let $P_M^*(\hat{x}^k)$ denote the set of minimizers of $J^k(P)$ over $\mathcal{P}^M(Q)$. For $k = 1, 2, \cdots$ and $M = 1, 2, \cdots$, let $\{P_M^{*k}\}$, $P_M^{*k} \in P_M^*(\hat{x}^k)$, be any sequence of minimizers in $\mathcal{P}(Q)$. By compactness there exists a convergent subsequence $\{P_{M_\ell}^{*k_j}\}$ such that

$$
\lim_{M_\ell, k_j \to \infty} P_{M_\ell}^{*k_j} = \tilde{P} \in \mathcal{P}(Q)
\tag{11}
$$

in the ρ metric.

First note that for any $P_{M_\ell} \in \mathcal{P}^{M_\ell}(Q)$

$$
J^{k_j}(P_{M_\ell}^{*k_j}) \leq J^{k_j}(P_{M_\ell}).
\tag{12}
$$

Then by definition of $J^{k_j}(P_{M_\ell}^{*k_j})$, (10), and (11)

$$
\begin{aligned}
\lim_{k_j, M_\ell \to \infty} J^{k_j}(P_{M_\ell}^{*k_j}) &= \lim_{k_j, M_\ell \to \infty} \sum_{i=1}^{n} |\mathcal{E}[x(t_i, q)|P_{M_\ell}^{*k_j}] - \hat{x}_i^{k_j}|^2 \\
&= \sum_{i=1}^{n} |\mathcal{E}[x(t_i, q)|\tilde{P}] - \hat{x}_i|^2 \\
&= J(\tilde{P}).
\end{aligned}
\tag{13}
$$

Let P be any element of $\mathcal{P}(Q)$. Since $\mathcal{P}_d(Q)$ is dense in $\mathcal{P}(Q)$ we can find a sequence $\{P_{M_\ell}\}$, $P_{M_\ell} \in \mathcal{P}^{M_\ell}(Q)$, so that $P_{M_\ell} \to P$ in the ρ metric as $M_\ell \to \infty$. Then by definition of $J^{k_j}(P_{M_\ell})$, and (10)

$$
\begin{aligned}
\lim_{k_j, M_\ell \to \infty} J^{k_j}(P_{M_\ell}) &= \lim_{k_j, M_\ell \to \infty} \sum_{i=1}^{n} |\mathcal{E}[x(t_i, q)|P_{M_\ell}] - \hat{x}_i^{k_j}|^2 \\
&= \sum_{i=1}^{n} |\mathcal{E}[x(t_i, q)|P] - \hat{x}_i|^2 \\
&= J(P).
\end{aligned}
\tag{14}
$$

So (12), (13), and (14) gives

$$
J(\tilde{P}) \le J(P)
$$

for any $P \in \mathcal{P}(Q)$. Hence $\tilde{P}$ is a minimizer of $J(P)$ over $P \in \mathcal{P}(Q)$, i.e., $\tilde{P} \in P^*(\hat{x})$.

Thus any sequence P_M^{*k} in $P_M^*(\hat{x}^k)$ has a subsequence $P_{M_\ell}^{*k_j}$ that converges to a $\tilde{P} \in P^*(\hat{x})$. So $dist(P_{M_\ell}^*(\hat{x}^{k_j}), P^*(\hat{x})) \to 0$ as $M_\ell \to \infty$ and $k_j \to \infty$. It follows that $dist(P_M^*(\hat{x}^k), P^*(\hat{x})) \to 0$ as $M \to \infty$ if $\hat{x}^k \to \hat{x}$. $\square$

In order to address computational issues, we use the family of approximate minimization problems defined above. If Q_d, $\mathcal{P}^M(Q)$, and $\mathcal{P}_d(Q)$ are defined as in Theorem 4.1, we know from Theorem 3.1 we can approximate any $P \in \mathcal{P}(Q)$ by $P_d \in \mathcal{P}_d(Q)$. Furthermore, from the results established above we also know we can approximate any $P \in \mathcal{P}(Q)$ by distributions $P_M \in \mathcal{P}^M(Q)$. By choosing M sufficiently large we obtain

$$
\int_Q x(t_i, q) \, dP(q) \approx \int_Q x(t_i, q) \sum_{j=1}^{M} \delta_{q_j^M}(q) \, dP(q) = \sum_{j=1}^{M} x(t_i, q_j^M) p_j.
$$

Thus we can approximate $J(q)$ by

$$
J^M(\overline{p}) \equiv \sum_{i=1}^{n} |\hat{x}_i - \sum_{j=1}^{M} x(t_i, q_j^M) p_j|^2
$$

9

where $\overline{p} = (p_1, p_2, \cdots, p_M)$ and $\sum_{j=1}^{M} p_j = 1$, $p_j \geq 0$, $p_j \in \mathcal{R}$, $1 \leq j \leq M$.

If we define

$$
\begin{aligned}
X_i^M &= (x(t_i, q_1^M), x(t_i, q_2^M), \ldots, x(t_i, q_M^M))^T, \; X^M = [X_1^M, \cdots, X_n^M] \\
\hat{X} &= (\hat{x}_1, \cdots, \hat{x}_n),
\end{aligned}
$$

then we can write

$$
J^M(\overline{p}) = \sum_{i=1}^{n} |\hat{x}_i - \overline{p} \cdot X_i^M|^2 = ||\hat{X} - \overline{p} \cdot X^M||_2^2.
$$

The approximate minimization problem thus reduces to a constrained optimization problem for a quadratic cost functional. Such problems are amenable to a number of standard algorithms.

Thus we see that any solution $\overline{p}$ must satisfy $\overline{p} \cdot X^M = \hat{X}$ and that if X^M is nonsingular, $\overline{p}$ is uniquely determined and depends continuously on the data $\hat{X}$.

5 Examples

We present a series of examples to illustrate a computational algorithm arising from the discussions of the previous sections. Each example uses the same system motivated by a problem in the assessment of the efficiency of a vaccination program [6, 2] by using data $\{\hat{x}_i\}$ for the aggregate population $x(t_i)$ of vaccinated but not infected individuals at time t_i. The evolution of the population is given by

$$
\dot{x}(t) = -qG(t)x(t), \; x(0) = x_0, \tag{15}
$$

where $G(t)$ represents the known rate of exposure to infection, q is the susceptibility to "environmental exposure" (a parameter to be chosen from an admissible parameter set Q) subsequent to vaccination of the population at time $t = 0$, and x_0 is the known number of individuals initially vaccinated.

All calculations were carried out using MATLAB routines. We let $Q = [0, 1]$ and define $Q_M = \{\frac{j-1}{M-1}\}_{j=1}^{M}$. Note $Q_d = \cup_{M=1}^{\infty} Q_M$ is a countable, dense subset of Q. For a given positive integer M, we would like to find a P^* that is a solution to (9) where $\mathcal{P}^M(Q)$ is defined as in (6) with our given Q_M. The data $\{\hat{x}_i\}$ is simulated data we generate.

To generate simulated data we start by taking N samples, $\{q_j^S\}_{j=1}^{N}$ on q^* from Q. The time interval $T = [0, 1]$ is discretized by $t_i = \frac{i}{n}$, $0 \leq i \leq n$, where $\frac{1}{n} = 0.01$. "Data" $\{\hat{x}_i\}$ is generated by first solving (15) at each point $x(t_i, q_j^S)$ with $G(t)$ given by

$$
G(t) = 130 \times \begin{cases} 0 & 0 \leq t \leq 0.1 \\ \frac{t-0.1}{0.8} & 0.1 < t \leq 0.9 \\ 1 & 0.9 < t \leq 1. \end{cases}
$$

10

Then, to average solutions from all samples, we define $x_i = \frac{1}{N} \sum_{j=1}^{N} x(t_i, q_j^S)$. Relative random noise was added to the solutions so that the "data" was given by $\hat{x}_i = x_i[1 + \epsilon_i]$ where the ϵ_i are independent Gaussian random variables with mean zero and variance σ^2.

In our first example, the samples of q^* are chosen from a normally distributed random variable on Q with mean 0.5 and standard deviation 0.1. The following figures display the optimal estimated discrete probability densities represented by $\overline{p} = (p_1, \cdots, p_M)$ and the corresponding probability distributions $\overline{P}_M = \sum_{j=1}^{M} p_j \delta_{q_j}$. In Figure 1 we present results of the optimization using data that was generated as described with 0% relative error. In Figure 2 the data was generated with 5% relative error, and in Figure 3 data was generated with 10% relative error. Note in each case as M increases the probability *distributions* are converging in the Prohorov metric as guaranteed by the theory, while the discrete *densities* do not converge in any sense on $Q = [0, 1]$. In each plot of the probability density, the "x"'s are the actual distribution of the generated data for q^* and in each plot of the probability distribution, the dashed line is the continuous distribution associated with q^* that the discrete distributions are attempting to approximate. The optimal distributions are graphed with piecewise constant solid lines.

In addition to testing the inverse problem on q^* with a Gaussian distribution, we also carried out the inverse problem with $M = 9$ and 0% and 5% relative error for the following q^* ($x(t_i; q^*)$ is again represented by a $1 \times M$ vector of values and the distribution for q^* is approximated by $\overline{p} = (p_1, \cdots, p_9)$):

- A delta function at 0.5, $q^* = 0.5$, with results presented in Figure 4;

- Two delta functions at 0.25 and 0.75, $q^*(j) = \left\{ \begin{array}{ll} 0.25 & 0 \leq j < N/2 \\ 0.75 & N/2 \leq j < 1 \end{array} \right.$, with results in Figure 5;

- Two skewed delta functions at 0.15 and 0.6, $q^*(j) = \left\{ \begin{array}{ll} 0.15 & 0 \leq j < N/3 \\ 0.6 & N/3 \leq j < 1 \end{array} \right.$, with results in Figure 6;

- A uniform distribution on $[0.25, 0.75]$, with results in Figure 7;

- A uniform distribution on $[0.1, 0.3] \cup [.55, .75]$, with results in Figure 8;

- A bimodal Gaussian distribution with one mean at 0.3 and the other at 0.7 and standard deviation 0.06, with results in Figure 9;

- A right skewed distribution with mean 0 and standard deviation 0.2, with results in Figure 10.

In each example the estimated probability distribution is a reasonable approximation of the continuous distribution, both with no error and with 5% relative error on the data.
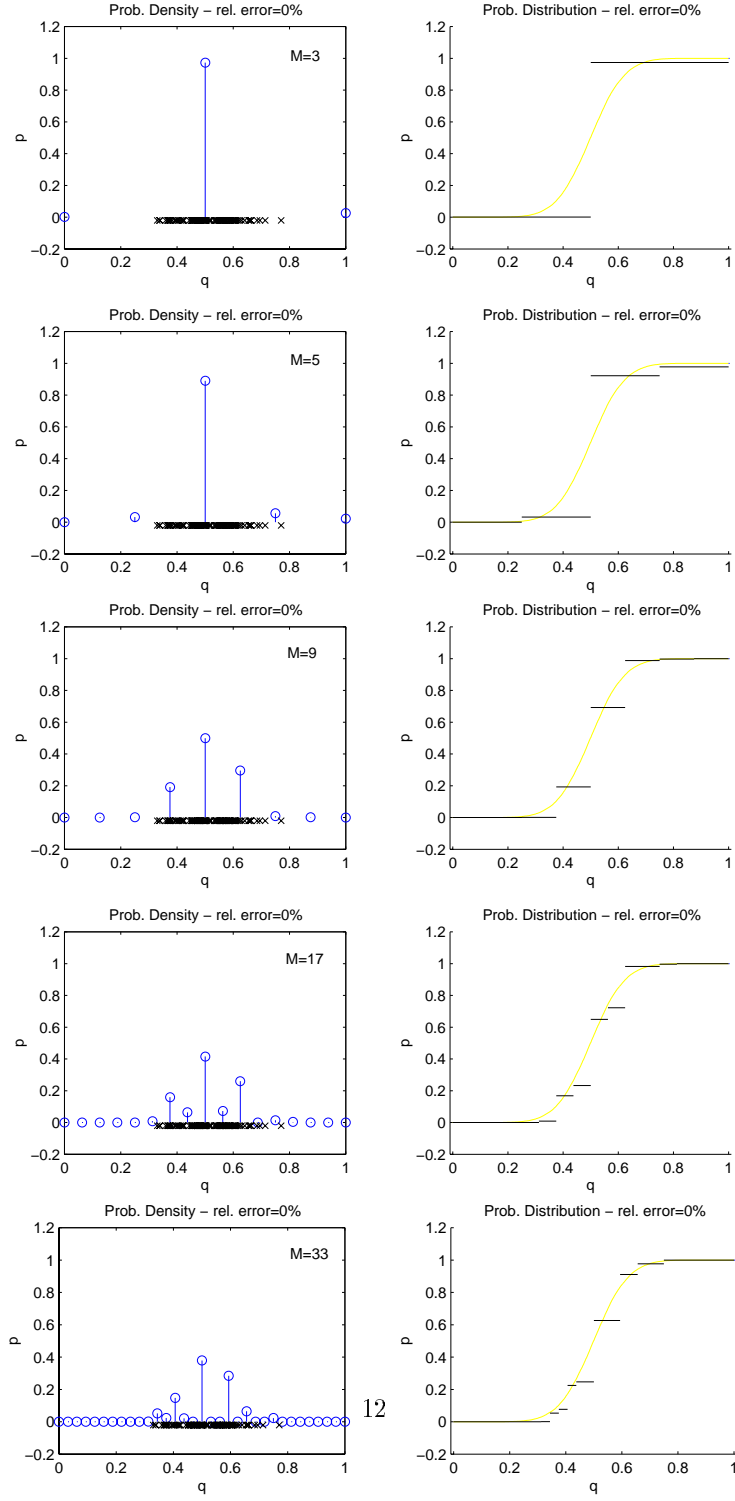
Figure 1: Approximate probability densities and probability distributions for q^* normally distributed and 0% error on data.
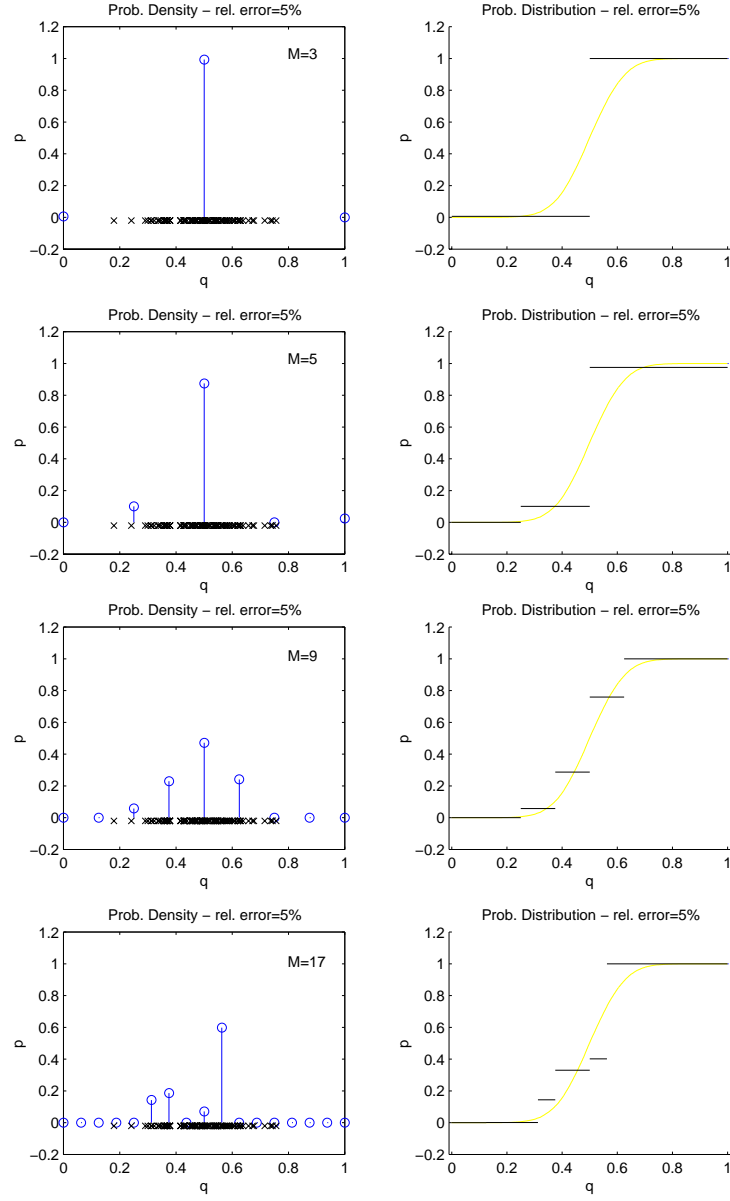
Figure 2: Approximate probability densities and probability distributions for q^* normally distributed and 5% error on data.
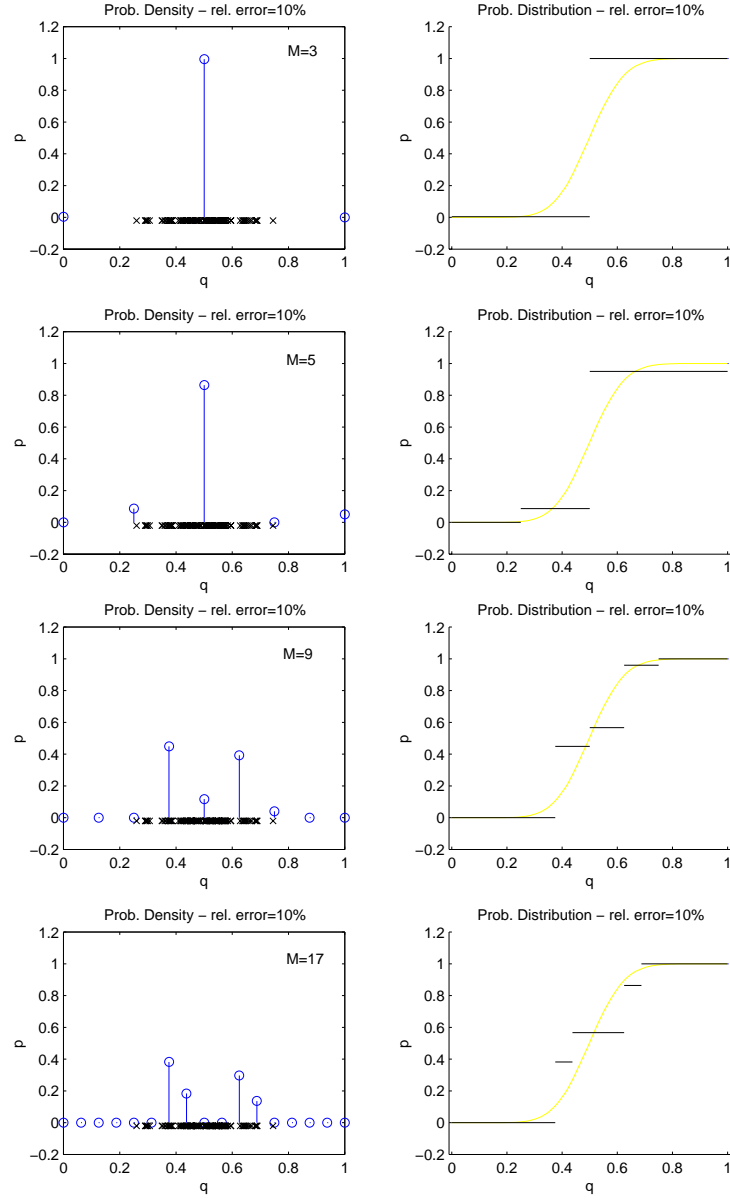
Figure 3: Approximate probability densities and probability distributions for q^* normally distributed and 10% error on data.
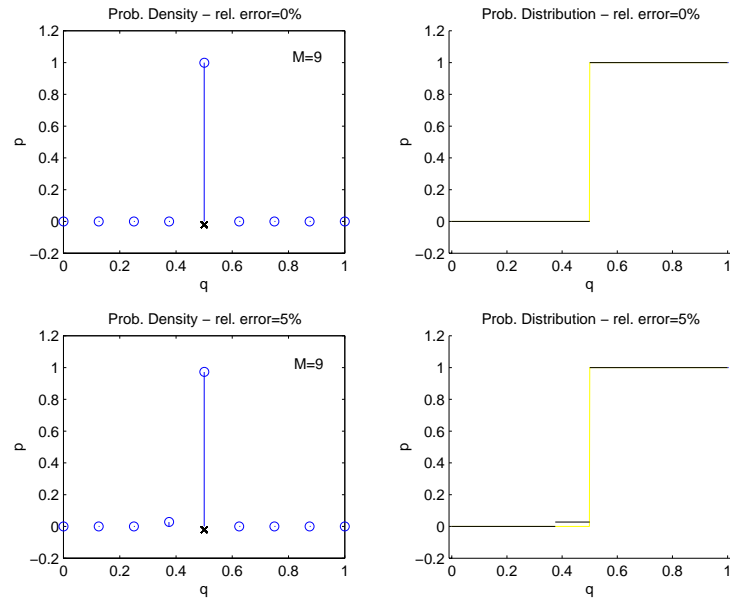
Figure 4: Approximate probability density and probability distribution for $q^* = 0.5$ with $M = 9$ and relative errors 0% and 5%.
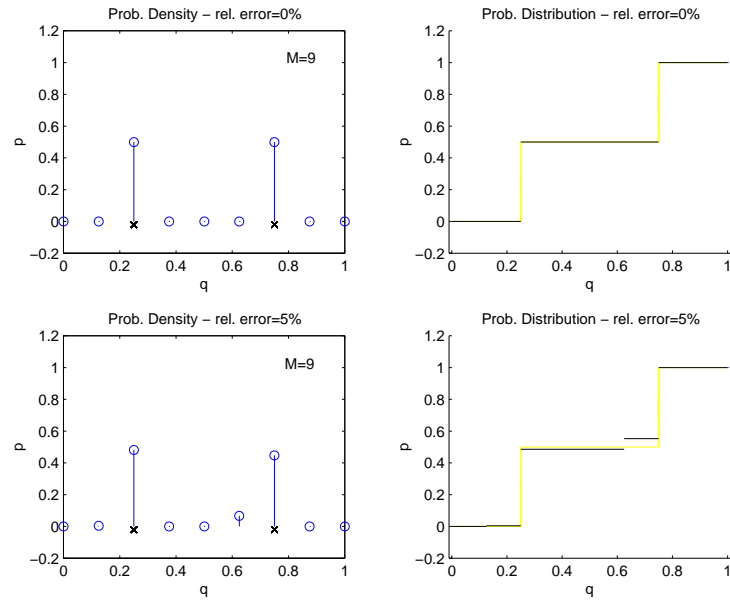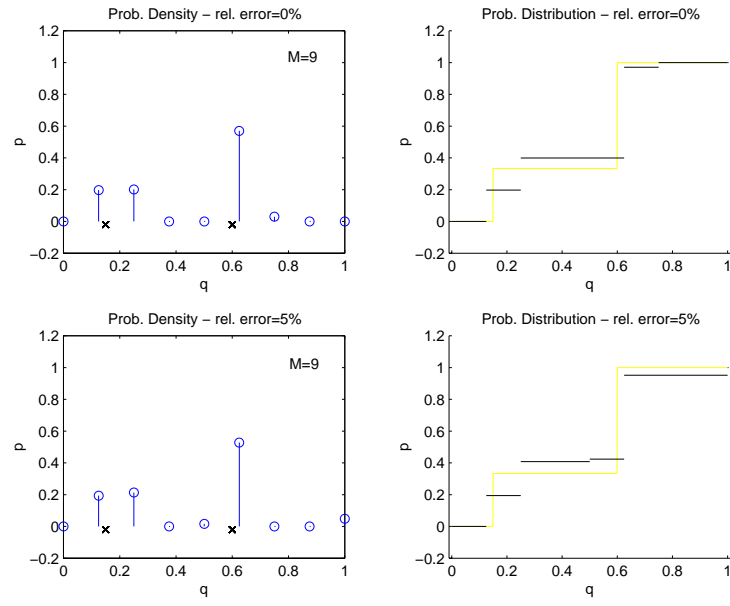
Figure 5: Approximate probability density and probability distribution for q^* two delta functions with $M = 9$ and relative errors 0% and 5%.

Figure 6: Approximate probability density and probability distribution for q^* two skewed delta functions with $M = 9$ and relative errors 0% and 5%.
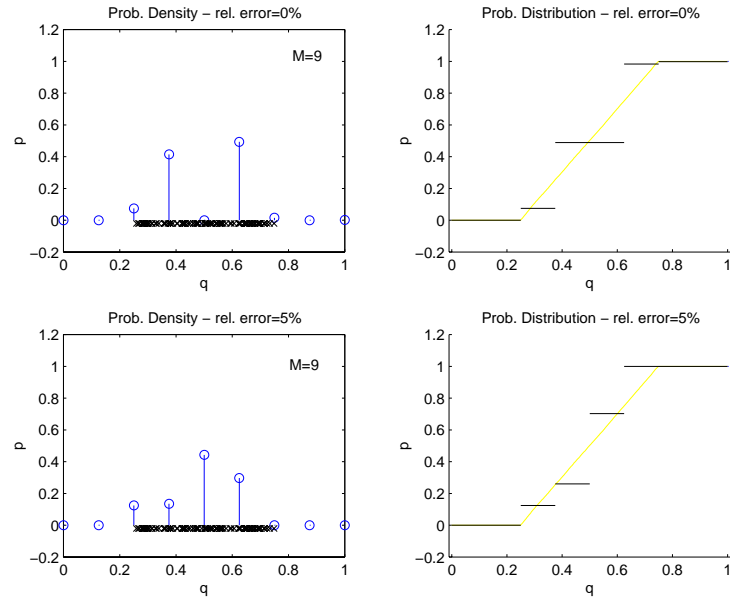
Figure 7: Approximate probability density and probability distribution for q^* uniformly distributed on $[.25, .75]$ with $M = 9$ and relative errors 0% and 5%.
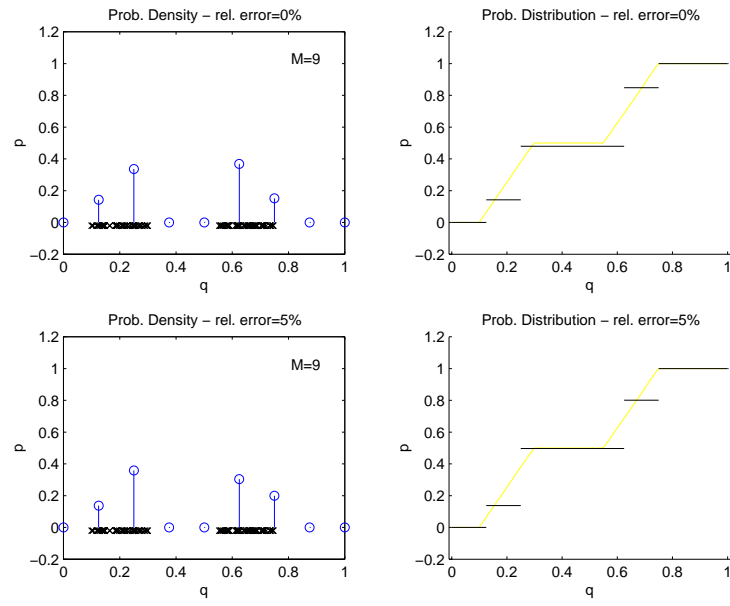
Figure 8: Approximate probability density and probability distribution for q^* uniformly distributed on $[.1, .3] \cup [.55, .75]$ with $M = 9$ and relative errors 0% and 5%.
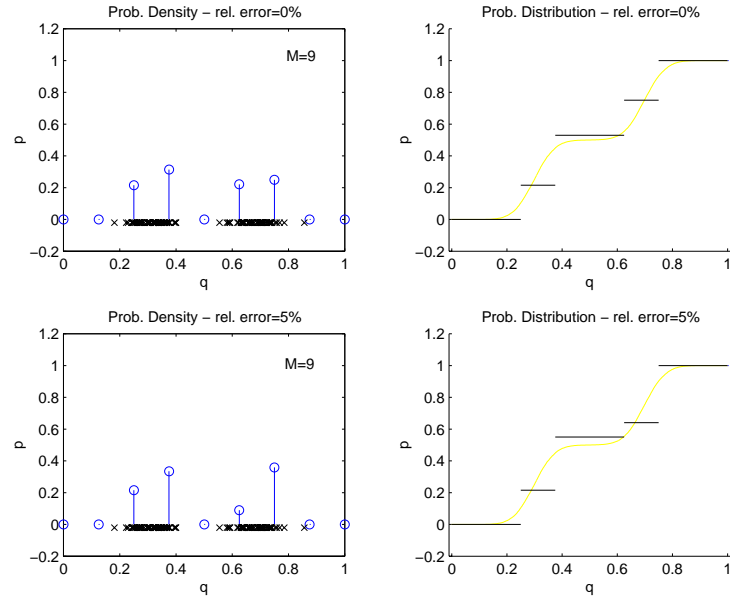
Figure 9: Approximate probability density and probability distribution for q^* binormally distributed with mean 0.3 and 0.7 and standard deviation 0.06 for each with $M = 9$ and relative errors 0% and 5%.
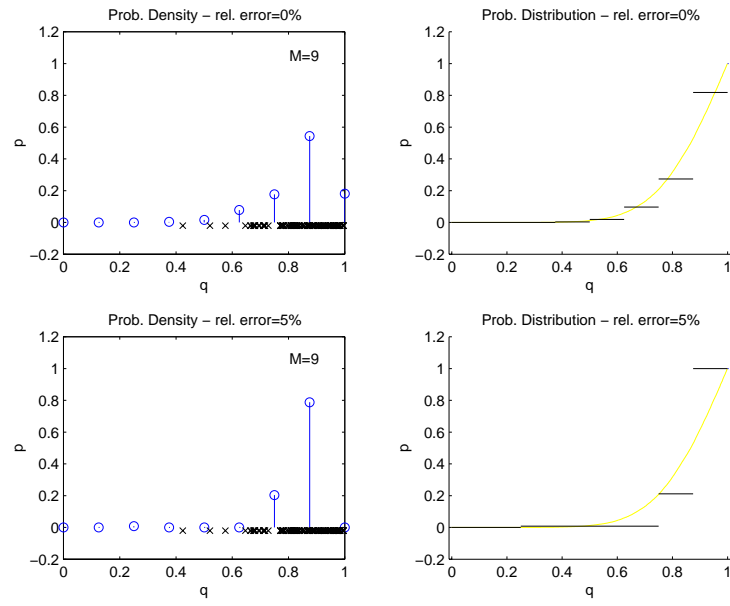
Figure 10: Approximate probability density and probability distribution for q^* right skew distributed with mean 0 and standard deviation0.2 with $M = 9$ and relative errors 0% and 5%.

6   Concluding Remarks

In the discussions above we have presented one approach to the quantification and computational treatment of uncertainty in inverse problems of a least squares formulation. By treating the estimated parameter as a random variable with unknown distribution, we conceptually reformulate the deterministic parameter estimation problem into a problem of estimation of a random variable using sampled data from a dynamical system which depends on the parameter. We use powerful but basic results from probability theory to develop a theoretical basis for these new problems. Approximation results along with continuous dependence of estimates on data and method stability are discussed. To test the ideas we present a series of numerical examples based on a single model arising in vaccination and susceptibility problems. The computational results presented support the efficacy of our approach and illustrate well the theoretical convergence results given in the paper.

References

[1] H.T. Banks, Remarks on Uncertainty Assessment and Management in Modeling and Computation, CRSC-TR98-39, November, 1998.

[2] H.T. Banks, B.G. Fitzpatrick, and Y. Zhang, "Estimation of Distributed Individual Rates from Aggregate Population Data", CRSC-TR94-13, Sept. 1994, in *Differential Equation and Applications to Biology and to Industry*, ed. by M. Martelli et al, World Scientific Press, 1996, pp.13-22.

[3] H.T. Banks and K. Kunisch. *Estimation Techniques for Distributed Parameter Systems*, Birkhäuser, Boston, 1989.

[4] H.T. Banks, R.C. Smith and Y. Wang. *Smart Material Structures Modeling, Estimation and Control*, Masson/J. Wiley, Paris/Chichester, 1996.

[5] P. Billingsley. *Convergence of Probability Measures*, Wiley, New York, 1968.

[6] R.C. Brunet, C.J. Struchiner and M.E. Halloran, "On the Distribution of Vaccine Protection Under Heterogeneous Response", *Math. Biosc.*, 116 (1993), pp. 111-125.